JJBS

Volume 16, Number 4,December 2023
ISSN 1995-6673
Pages 611 – 619
https://doi.org/10.54319/jjbs/160406

Jordan Journal of Biological Sciences

Comprehensive Characterization and Expression Profiling of the GATA Transcription Factor in Sugar Beet (*Beta vulgaris* L.) Suggests Their Potential Roles in Taproot Development and Biotic Stress Response

Le Thi Man[1], Phi Bang Cao[1], Thi Thanh Huyen Tran[2], Nguyen Quoc Trung[3], Dong Huy Gioi[3], Ninh Thi Thao[3], Quynh Thi Ngoc Le[4], Hong Viet La[5], Thi Xuan Quyen Vu[1], Ha Duc Chu[6,*]

[1]*Faculty of Natural Sciences, Hung Vuong University, Phu Tho Province 35000, Vietnam;* [2]*Faculty of Biology, Hanoi National University of Education, Xuan Thuy Road, Cau Giay District, Hanoi City 122300, Vietnam;* [3]*Faculty of Biotechnology, Vietnam National University of Agriculture, Gia Lam District, Hanoi City 122300, Vietnam;* [4]*Department of Biotechnology, Thuyloi University, Tay Son Street, Dong Da District, Hanoi City 122300, Vietnam;* [5]*Institute of Research and Application, Hanoi Pedagogical University 2, Phuc Yen City, Vinh Phuc Province 280000, Vietnam;*[6]*Faculty of Agricultural Technology, University of Engineering and Technology, Vietnam National University Hanoi, Xuan Thuy Road, Cau Giay District, Hanoi City 122300, Vietnam;*

Received: November 29, 2022; Revised: March 31, 2023; Accepted: April 3, 2023

Abstract

GATA transcription factors (TFs) are well-characterized as major regulators whose DNA-binding domain is a type IV zinc finger motif. In this present study, we identified and characterized the GATA TFs in sugar beet (*Beta vulgaris*). A total of 16 BvGATA TFs from sugar beet has been reported. Based on the web-based tools, our analysis indicated that the BvGATA TFs exhibited a high variation in their physic-chemical features and gene structure. Four segmental duplication events that occurred in the *BvGATA* gene family have been predicted. The phylogeny analysis demonstrated that the BvGATA TFs could be classified into four clades. Interestingly, the transcriptional changes of the *BvGATA* genes were analyzed according to three available transcriptome databases. We found that a majority of the *BvGATA* genes showed significant transcriptional changes in major tissues under adverse environmental conditions. To sum up, our findings could provide a cornerstone to deeply understand the GATA TFs in sugar beet.

Keywords: Characterization, expression profiles, GATA, identification, sugar beet, transcription factor.

1. Introduction

Sugar beet (*Beta vulgaris* L.) has been regarded as the major sugar-yielding crop that is cultivated commercially in the world. Providing 30 - 35% of annual sugar production in the world (Monteiro *et al.*, 2018), sugar beet was reported to be the second-largest source of sugar (Zhang *et al.*, 2016). This crop could be used as human food and cattle feed (Evans and Messerschmidt, 2017; Monteiro *et al.*, 2018; Zhang *et al.*, 2016) and raw materials for bioethanol production (Mall *et al*., 2021; Pavlečić *et al*., 2017). Interestingly, containing various natural pigments (Chhikara *et al.*, 2019), sugar beet is believed to offer beneficial physiological effects as a functional food (Chen *et al.*, 2021; Mirmiran *et al.*, 2020). It, accordingly, would be more significant to investigate the growth and development processes in this important sugar-producing crop on a molecular basis.

In higher plant species, transcription factors (TFs) could play crucial roles in various biological pathways, particularly related to unfavorable condition stimulate (Baillo *et al.*, 2019; Fujita et al., 2011). Among them, GATA, a group of highly conserved type IV zincs finger motifs is well-characterized as one of the general eukaryotic-specific TFs (Schwechheimer *et al.*, 2022). Structurally, GATA TFs include a single domain that specifically binds to a conserved DNA motif, like -WGATAR- in the promoter regions (Behringer and Schwechheimer, 2015; Schwechheimer *et al.*, 2022; Teakle *et al.*, 2002). To date, the GATA TFs have been identified in a variety of higher plant species, including dicotyledonous plants, like *Arabidopsis thaliana* (Teakle *et al.*, 2002), soybean (*Glycine max*) (Zhang *et al.*, 2015), apple (*Malus domestica*) (Chen *et al.*, 2017), grape (*Vitis vinifera*) (Zhang *et al.*, 2018), cotton (*Gossypium* spp.) (Zhang *et al.*, 2019), chickpea (*Cicer arietinum*) (Niu *et al.*, 2020), pepper (*Capsicum annuum*) (Yu *et al.*, 2021), cucumber (*Cucumis sativus*) (Zhang *et al*., 2021), potato (*Solanum tuberosum*) (Yu *et al*., 2021) and Rosaceae species (Manzoor *et al*., 2021), and monocotyledonous plants, like rice (*Oryza sativa*) (Reyes *et al*., 2004), purple false brome (*Brachypodium distachyon*) (Peng *et al*., 2021), *Populus* spp. (Kim *et al*., 2021) and wheat

* Corresponding author. e-mail: cd.ha@vnu.edu.vn.

(*Triticum aestivum*) (Feng *et al.*, 2022). Unfortunately, the GATA TFs in sugar beet remain poorly understood, even though the genome of this important crop has been published recently (Dohm *et al.*, 2014).

In this current study, we performed a systematic analysis of the GATA TFs in sugar beet by using computational approaches. We first screened and identified all putative members of the GATA TFs in sugar beet. The general characteristics of the GATA TFs in sugar beet have been analyzed. Additionally, the gene duplication events and gene structures have been predicted in the GATA TFs in sugar beet. Finally, we carried out a re-analysis of the expression profiles of the *GATA* genes under numerous conditions in sugar beet.

2. Materials and Methods

2.1. 2.1. Database searches for the GATA TFs in sugar beet

The GATA TFs in sugar beet were searched by using various well-known databases. Briefly, the PlantTFDB v4.0 (Jin *et al.*, 2017) was firstly used to screen all putative GATA amino acid (aa) sequences in the sugar beet proteome (Dohm *et al.*, 2014) as previously reported (Chu *et al.*, 2018; La *et al.*, 2022; Niu *et al.*, 2020). The Pfam database (Mistry *et al.*, 2021) was then employed to confirm the existence of the GATA TF conserved domains as previously described (Schwechheimer *et al.*, 2022). The GATA TFs were annotated by BlastP-ing against the sugar beet assemblies (Dohm *et al.*, 2014) available from the NCBI and Phytozome v12.0 (Goodstein *et al.*, 2012) as previously guided (La *et al.*, 2022).

2.2. 2.2. Estimation of protein features of the GATA TFs in sugar beet

The aa sequences of the GATA TFs were used to analyze the general features of proteins as previously reported (Chu *et al.*, 2018; La *et al.*, 2022; Niu *et al.*, 2020). Particularly, five properties, including protein size (aa residues), protein mass (kilo Dalton, kDa), isoelectric point (pI), aliphatic index (AI) and grand average of hydropathicity (GRAVY) were calculated by the ExPaSy tool (Gasteiger *et al.*, 2003; Gasteiger *et al.*, 2005).

2.3. 2.3. Prediction of the subcellular localization of the GATA TFs in sugar beet

The full-length aa sequences of the GATA TFs in sugar beet were used as seed sequences to query in the YLOC tool (Briesemeister *et al.*, 2010) as previously described (La *et al.*, 2022; Niu *et al.*, 2020). Organelle-specific signal peptides were screened in the full-length aa sequences of each protein. Ten major organelles were investigated for the plant model as previously reported (Briesemeister *et al.*, 2010).

2.4. 2.4. Generation of phylogenetic tree of the GATA TFs in sugar beet

To generate the phylogenetic tree of the GATA TFs in sugar beet, full-length aa sequences were analyzed in various tools as previously described (La *et al.*, 2022). Firstly, ClustalX 2.0 software (Larkin *et al.*, 2007) was used to align the full-length aa sequences. An unrooted phylogenetic tree of the GATA TFs in sugar beet was then generated by using the Maximum likelihood (ML) estimation with 1000 times bootstrap and other default parameters in the MEGA tool (Kumar *et al.*, 2016). We also collected 29 well-established members of the GATA TFs in *A. thaliana* from the previous study (Teakle *et al.*, 2002) for other ML- phylogenetic tree by the MEGA tool. All phylogenetic trees were then visualized by using the Adobe Illustrator software.

2.5. 2.5. Investigation of gene features of the GATA TFs in sugar beet

The structural information on the GATA TFs in sugar beet was analyzed as previously described (La *et al.*, 2022; Niu *et al.*, 2020). Briefly, the full-length nucleotide sequences of each gene encoding the GATA TFs obtained in the previous *in silico* analyses were used to apply in the Gene Structure Display Server (GSDS) website (Hu *et al.*, 2015). The order of exon/intron structures of genes encoding the GATA TFs was then constructed based on the ML phylogenetic tree (Kumar *et al.*, 2016). To carry out enrichment analysis on gene sets, the aa sequences of the GATA TFs were analyzed by using the Blast2GO software (Conesa *et al.*, 2005)

2.6. 2.6. Prediction of gene duplication of the GATA TFs in sugar beet

To analyze the gene duplication in the *GATA* gene family in sugar beet, the full-length nucleotide sequences were used as templates for *in silico* prediction by following the previous studies (La *et al.*, 2022; Niu *et al.*, 2020) with minor adjustment. Particularly, the identity matrix of all aligned nucleotide sequences of *GATA* genes was constructed by the BioEDIT software (Hall, 1999). All duplication events were then applied in the DNASp tool (Rozas *et al.*, 2017) to calculate the ratio between non-synonymous substitutions per non-synonymous site (Ka) and synonymous substitutions per synonymous site (Ks) as previously described (La *et al.*, 2022; Niu *et al.*, 2020).

2.7. 2.7. Analysis of the expression profiles of the GATA TFs in sugar beet

To investigate the transcriptional changes of the *GATA* TFs, we performed a re-analysis of four transcriptome datasets available from the NCBI Gene expression omnibus (NCBI GEO) (Barrett *et al.*, 2013). Particularly, GSE107627 provided a dataset related short-term and long-term treatment of alkaline solution in leaves (Zou *et al.*, 2020). We also analyzed GSE114968 dataset, which provided expression data of salt- treated (1 day and 7 days) roots at seedling stages (Liu *et al.*, 2020). Finally, we re-analyzed the third microarray dataset (GSE135555) related to the beet cyst nematode inoculation in roots (Ghaemi *et al.*, 2020). Responsive gene was defined by a fold-change cut-off (fold-change $\geq$ 1.5-fold or $\leq$ -1.5-fold).

3. Results and Discussion

3.1. Identification and annotation of the GATA TFs in sugar beet

To find all members of the GATA TFs in sugar beet, we explored the PlantTFDB (Jin *et al.*, 2017) and the newest sugar beet assembly (Dohm *et al.*, 2014) from NCBI and Phytozome (Goodstein *et al.*, 2012). After validating by the Pfam (Mistry *et al.*, 2021), we explored a total of 16 conserved GATA aa sequences in the sugar beet

proteome (Table 1). The annotations, like ProteinID and locusID, were accordingly retrieved from the NCBI and Phytozome databases (Goodstein *et al.*, 2012), respectively and showed in Table 1. Finally, according to the physical location on genome, we assigned whole 16 identified GATA aa sequences from BvGATA01 to BvGATA16 (Table 1).

Table 1. Summary of the BvGATA TFs in sugar beet

#	Gene	Phytozome locus	NCBI Protein ID	Gene size	Protein size	MW (kDa)	pI	AI	GRAVY	SCL
1	BvGATA01	EL10Ac1g00497	XP_010671518.1	5793	297	32.88	5.94	62.29	-0.74	Nucleus
2	BvGATA02	EL10Ac1g00493	XP_010671467.1	6631	346	37.72	5.74	69.05	-0.67	Nucleus
3	BvGATA03	EL10Ac2g03883	XP_010667257.2	3488	318	35.72	8.94	60.19	-0.85	Cytoplasm
4	BvGATA04	EL10Ac2g04003	XP_010668515.1	1794	375	41.53	6.52	60.85	-0.67	Cytoplasm
5	BvGATA05	EL10Ac3g04929	XP_010670911.1	667	155	16.7	10.02	64.84	-0.82	Cytoplasm
6	BvGATA06	EL10Ac4g09843	XP_010674149.1	580	146	16.76	9.26	48.08	-0.94	Nucleus
7	BvGATA07	EL10Ac5g12284	XP_010677131.2	6317	548	61.06	5.72	65.84	-0.66	Nucleus
8	BvGATA08	EL10Ac7g16992	XP_010683872.2	2977	308	33.88	8.41	65.58	-0.54	Cytoplasm
9	BvGATA09	EL10Ac6g14415	XP_010681549.2	3862	491	55.57	8.89	57.56	-0.93	Cytoplasm
10	BvGATA10	EL10Ac7g17612	XP_010684625.1	673	138	14.99	9.56	70.00	-0.6	Cytoplasm
11	BvGATA11	EL10Ac7g18179	XP_010696086.2	10707	299	32.89	5.94	55.75	-0.82	Nucleus
12	BvGATA12	EL10Ac9g21650	XP_010690771.1	2553	320	35.22	8.36	58.78	-0.54	Cytoplasm
13	BvGATA13	EL10Ac9g22059	XP_010691384.1	10336	353	38.85	4.97	66.86	-0.60	Nucleus
14	BvGATA14	EL10Ac9g22099	XP_010691426.1	1228	309	33.85	6.43	47.96	-0.75	Cytoplasm
15	BvGATA15	EL10As8g23860	XP_010679831.2	957	281	31.63	8.73	46.16	-0.93	Cytoplasm
16	BvGATA16	EL10Ac6g15556	XP_010679730.1	1203	259	28.39	7.61	35.48	-0.92	Cytoplasm

Note: -: No information, AI: Aliphatic index, pI: Iso-electric point, GRAVY: Grand average of hydropathicity, SCL: Sub-cellular localization

Recently, great efforts have been recorded in order to identify and intensively analyze the GATA TFs in numerous plant species. The numbers of the GATA TFs in dicotyledonous plants were varied from 19 (in grape) (Zhang *et al.*, 2018) to 64 members (in soybean) (Zhang *et al.*, 2015). Another case in cotton species, a total of 46, 46 and 87 members of the GATA TFs has been reported in *G. arboreum*, *G. raimondii* and *G. hirsutum* (Zhang *et al.*, 2019). In four Rosaceae species, 92 members of the GATA TFs were found, with 32 members from *Pyrus bretschneideri*, 18 members from *Prunus avium*, 20 members from *Prunus mume* and 22 members from *Prunus persica* (Manzoor *et al.*, 2021). In monocotyledonous plants, the amounts of the GATA TFs were also variable. For example, the GATA TFs in rice have been reported to contain 28 members (Reyes *et al.*, 2004), as similar as the members in purple false brome (Peng *et al.*, 2021). Additionally, 79 members of the GATA TFs have been investigated in wheat (Feng *et al.*, 2022). To sum up, 16 members of the BvGATA TFs found throughout the whole genome of sugar beet is close to which is close to the 18, 19 and 20 GATA TFs identified in grape (Zhang *et al.*, 2018), *P. avium* and *P. mume* (Manzoor *et al.*, 2021), respectively, but greatly less than the GATA TFs identified in other dicots, like *P. persica* (22 members) (Manzoor *et al.*, 2021), chickpea (25 members) (Niu *et al.*, 2020), *A. thaliana* (29 members) (Teakle *et al.*, 2002), *P. bretschneideri* (32 members) (Manzoor *et al.*, 2021), apple (35 members) (Chen *et al.*, 2017), seven *Populus* spp. (33 to 40 members) (Kim *et al.*, 2021), *G. arboreum* (46 members) (Zhang *et al.*, 2019), *G. raimondii* (46 members) (Z. Zhang *et al.*, 2019), potato (49 members) (Yu *et al.*, 2021), soybean (64 members) (C. Zhang *et al.*, 2015) and *G. hirsutum* (87 members) (Zhang *et al.*, 2019). Our comparisons revealed that the number of GATA TF members varies greatly between higher plant species.

3.2. Analysis of the physical and chemical features of the GATA TFs in sugar beet

In this study, the ExPaSy Protparam tool (Gasteiger *et al.*, 2003; Gasteiger *et al.*, 2005) was applied to analyze full-length aa sequences of 16 BvGATA members in sugar beet. As a result, the physical and chemical properties of the BvGATA TFs in sugar beet were provided in Table 1. Particularly, the amounts of aa residues encoded by 16 BvGATA TFs ranged from 138 (BvGATA10) to 548 aa residues (BvGATA07), with an average of nearly 309 aa (Table 1). Next, the weights of BvGATA TFs were distributed between 14.99 (BvGATA10) and 61.06 kDa (BvGATA07), with an average of approximately 34.23 kDa (Table 1). The theoretical pI scores of the BvGATA TFs were recorded to be acidic, ranging from 4.97 (BvGATA13) to 6.52 (BvGATA04) and base, ranging from 7.61 (BvGATA16) to 10.02 (BvGATA05), with an average of approximately 7.56 (Table 1). Additionally, the AI values of the BvGATA TFs were varied from 35.48 (BvGATA16) to 70.00 (BvGATA10), with an average of nearly 58.45 (Table 1). Finally, the GRAVY scores of all members of the BvGATA TFs in sugar beet were < 0, with average of approximately -0.75 (Table 1), suggesting that BvGATA TFs belonged to hydrophilic proteins.

Previously, the general characteristics of the GATA TFs were also comprehensively analyzed and reported in other higher plant species. The sizes of the GATA TFs in legumes, like soybean and chickpea ranged from 80 to 551 aa residues (9.1 to 60.8 kDa) (Zhang *et al.*, 2015) and from 133 to 541 aa residues (14.9 to 60.2 kDa) (Niu *et al.*, 2020), respectively, while the protein lengths of the GATA

TFs in grape were varied from 109 to 386 aa residues (Zhang *et al.*, 2018). The aa residues of the GATA TFs in potato were 118 and the largest was 380 (13.15 to 60.63 kDa) (R. Yu *et al.*, 2021), while the numbers of aa residues of the GATA TFs in apple were varied from 90 to 1161 (9.9 to 129.74 kDa) (Chen *et al*., 2017). In cotton species, the predicted GATA sequences consisted of 119 to 584 aa residues, with an average of 306 aa residues (Zhang *et al.*, 2019). Next, the pI values of the GATA TFs in higher plant species were confirmed to range from acidic to base. For example, the pI scores of the GATA TFs in soybean and chickpea have been reported to range from 4.63 to 9.66 (Zhang *et al*., 2015) and 4.27 to 10.27 (Niu *et al.*, 2020), respectively. Interestingly, all members of the GATA TFs in apple and chickpea were reportedly < 0 (Chen *et al.*, 2017; Niu *et al.*, 2020), suggested that the GATA TFs in these plants, or perhaps in other plant species were hydrophilic (Schwechheimer *et al*., 2022). To sum up, our results indicated that the BvGATA TFs in sugar beet, perhaps in many plant species showed high variation in their physical and chemical properties.

3.3. Subcellular localization and gene ontology analysis of the GATA TFs in sugar beet

The determination of the subcellular localization of proteins may suggest their potential function in the cellular metabolism (Goodin, 2018). Here, we predicted the subcellular localization of the BgGATA TFs in sugar beet by using the YLOC tool (Briesemeister *et al.*, 2010). As a result, a large amount of the BvGATA TFs (10 out of 16) was localized in the cytoplasm, while the remaining members of the BvGATA TFs (six out of 16) were predicted to distribute on the nucleus (Table 1).

Furthermore, the potential functions of the BvGATA TFs were annotated by using the gene ontology (GO) annotation analysis in sugar beet. As a result, 16 members of the BvGATA TFs in sugar beet were classified into three ontologies, like molecular function, biological process and subcellular localization. These BvGATA TFs were assigned with 15 GO terms belonging to the molecular function (Figure 1). Under the molecular function category, all BvGATA proteins were confirmed to act as TFs (Figure 1). Under the biological process category, all BvGATAs were involved in regulation of biological processes and 15 out of 16 members of the BvGATA TFs were predicted to play a role in response to stimulus (GO:0050896), regulation of nitrogen compound metabolic process (GO:0051171), regulation of metabolic process (GO:0019222) and regulation of cellular metabolic process (GO:0031323) (Figure 1). The GO analysis also indicated that all BvGATA TFs were localized in the nucleus, which is also confirmed by the YLoc prediction (Briesemeister *et al*., 2010).

These findings were also confirmed in the previous reports. Particularly, a member of the GATA TFs in purple false brome, namely in BdGATA13, was investigated to localize in the nucleus by green fluorescent protein tagging method (Peng *et al.*, 2021). Previously, the green fluorescent protein::GmGATA58 (a member of the GATA TFs in soybean) fusion protein driven by the CaMV 35S promoter exhibited a strong green fluorescent signal in the nucleus (Zhang *et al.*, 2020). It suggested that the GmGATA58 protein was indeed localized in nucleus (Zhang *et al.*, 2020). Moreover, the Gene Ontology annotation analysis predicted that a majority of the GATA TFs in four Rosaceae species anticipated their function into nuclear (Manzoor *et al.*, 2021).

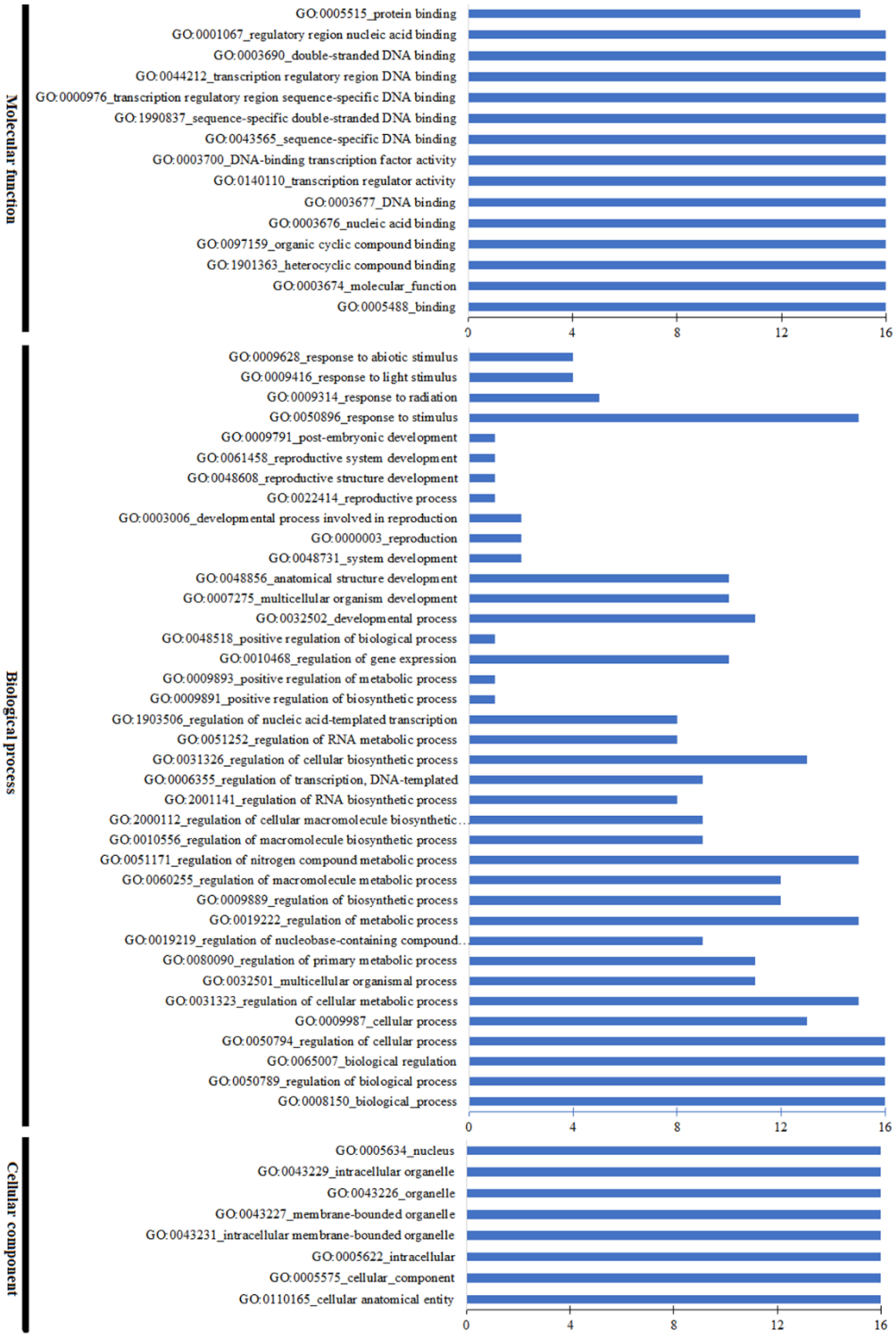

Figure 1. GO analysis involving in molecular function, biological processes, and cellular component of BvGATA investigated by NETGO 2.0

3.4. Phylogenetic analysis of the GATA TFs in sugar beet

In order to elucidate the phylogenetic relationships of the BvGATA TFs in sugar beet, an unrooted phylogenetic tree of whole 16 BvGATA TFs and well-characterized GATA TFs from *A. thaliana* (Teakle *et al.*, 2002) has been built. Based on the ML estimation, 16 BvGATA proteins were classified into four different sub-groups (Figure 2). Particularly, clade 1 contained six (out of 16) members, including BvGATA04, 09, 10, 11, 14 and 16 (Figure 2). Clade 2 and 3 had four (out of 16), like BvGATA01, 02, 12 and 15, and five (out of 16), members, like BvGATA03, 05, 06, 08 and 12, respectively (Figure 2). Finally, only one (out of 16) member of the BvGATA TFs, like BvGATA07 was distributed in the clade 4 (Figure 2). Phylogenetic tree strongly indicated that the categorization of the BvGATA TFs in sugar beet exhibited a similar trend when comparing with *A. thaliana*.

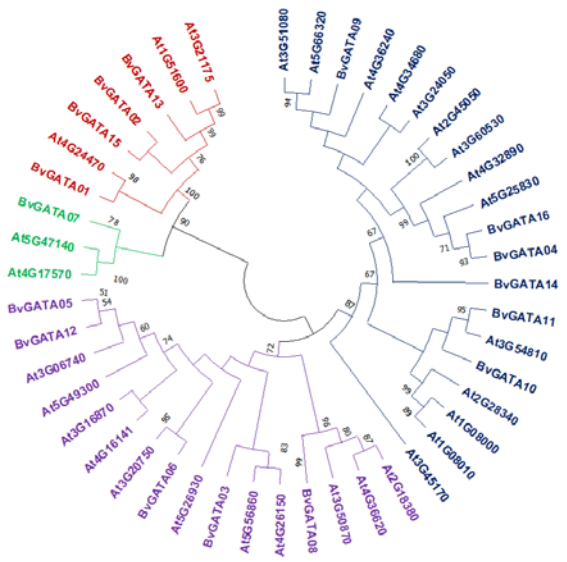

Figure 2. Phylogenetic tree of GATA family from sugar beet (Bv) and Arabidopsis (At).

Previously, the classification of the GATA TFs has been demonstrated in various higher plant species. Briefly, an unrooted phylogenetic tree between all members of the GATA TFs from *A. thaliana* (Teakle *et al*., 2002) and four Rosaceae species revealed that these GATA TFs were also classified into four clades (Manzoor *et al*., 2021). This result was confirmed in the report of the GATA TFs in *Populus* spp. (Kim *et al*., 2021). Particularly, a phylogenetic tree of the GATA TFs from *Populus* spp. and *A. thaliana* (Teakle *et al*., 2002) has been built in order to reveal sub-families (Kim *et al*., 2021). As expected, four sub-families have been identified; particularly clade I and IV contained the largest and smallest number of the GATA TFs from *Populus* spp., respectively (Kim *et al*., 2021). This phenomenon was confirmed in the GATA TFs in other plant species, like A. *thaliana* (Teakle *et al*., 2002), soybean (Zhang *et al*., 2015) and grape (Zhang *et al*., 2018).

3.5. Physical distribution, gene organization and gene duplication of the GATA TFs in sugar beet

In this study, we examined the chromosomal distributions of the 16 *BvGATA* genes. The *BvGATA* gene family was found to randomly localize on the sugar beet genome. For example, two members of the *BvGATA* gene family were located in chromosome 1 (Figure 3). Chromosome 2 also contained two *BvGATA* genes, while only one member from the *BvGATA* gene family has been found in each of chromosome 3, 4 and 5 (Figure 3). Next, three *BvGATA* genes were mapped on each of chromosome 6, 7 and 9 (Figure 3). Additionally, no *BvGATA* gene was found in chromosome 8 (Figure 3).

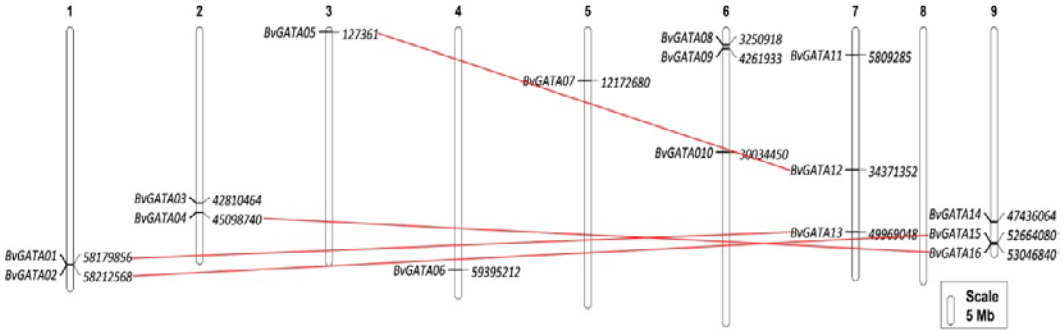

Figure 3. The chromosomal distribution of *BvGATA* genes in the sugar beet genome. The red lines indicated the duplication events. The chromosome number is indicated to the above of each chromosome

Next, we analyzed the exon/intron configuration of the *BvGATA* genes in sugar beet. Our results revealed the variable features of the four sub-groups (Figure 4). Particularly, five (out of six) members in clade 1 contained two exons, whereas only *BvGATA14* in clade 1 had three exons (Figure 4). Similarly, a majority member from clade 3 had three exons and only two (out of five) members in this clade, particularly *BvGATA06* and *08* contained two exons. Interestingly, *BvGATA01* and *13* from clade 2 had seven exons, while *BvGATA02* and *15* from clade 2 contained 10 exons (Figure 4). *BvGATA07* from clade 4 contained eight exons (Figure 4).

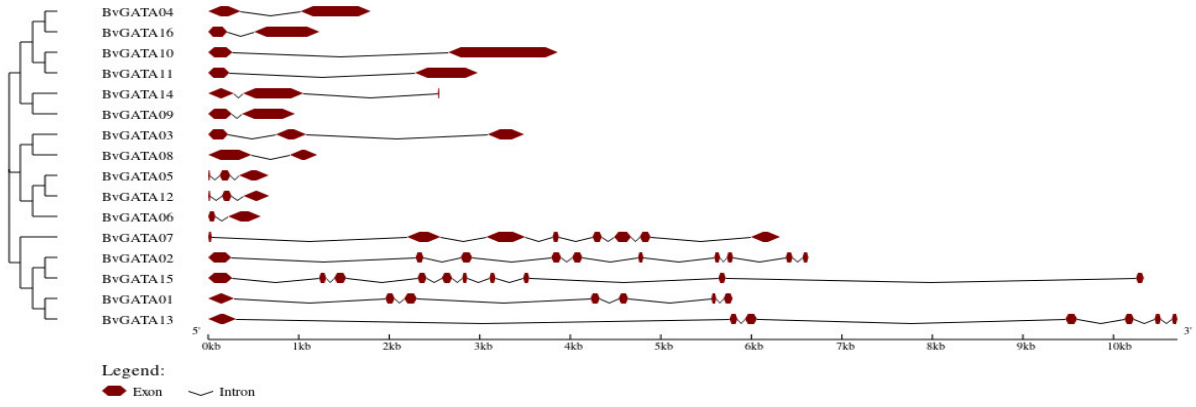

Figure 4. Gene exon/intron organizations of the *GATA* TF family in sugar beet

As an interesting part of this study, we also predicted the duplication events occurred in the *BvGATA* gene family in sugar beet by using various tools (Hall, 1999; Rozas *et al.*, 2017) as previously described (La *et al.*, 2022; Niu *et al.*, 2020). As well-described in Figure 3 and Table 2, a total of four duplicated genes has been reported in the *BvGATA* family. The similar levels at the nucleotide scale of whole events were varied from 50.0 (*BvGATA02* and *15*) to 57.5% (*BvGATA05* and *12*) (Table 2). Interestingly, we realized that all duplicated genes were localized in different chromosomes of the sugar beet genome (Figure 3). This finding revealed that the segmental duplication events were major reasons for the expansion of the *BvGATA* gene family. Furthermore, the rates of Ka/Ks of four duplicated genes were recorded to be less than 0, ranging from 0.36 (*BvGATA01* and *13*) to 0.61 (*BvGATA05* and *12*) (Table 2).

Table 2. Prediction of the duplication events in the *BvGATA* gene family in sugar beet

Duplicated genes	Mechanism	Similar level (%)	Ka	Ks	Ka/Ks
BvGATA04 and *16*	Segmental duplication	53.7	0.3760	0.6451	0.58
BvGATA05 and *12*	Segmental duplication	57.5	0.3913	0.6446	0.61
BvGATA01 and *13*	Segmental duplication	52.9	0.4880	1.3361	0.36
BvGATA02 and *15*	Segmental duplication	50.0	0.5717	1.5408	0.37

3.6. Expression patterns of the sugar beet GATA TFs under abiotic and biotic stress

In this study, of our interest, we explored the expression patterns of the *BvGATA* genes in various organs/tissues under different conditions. As a result, a heatmap was constructed by using R script. We realized that all *BvGATA* genes showed their variable expression in various main tissues under different conditions.

Under the alkaline treatment, a large number of the *BvGATA* genes has been reported to differentially express in leaves. Two genes (out of 16), like *BvGATA02* and *07* were up-regulated in treated leaves, whereas eight genes (out of 16), including *BvGATA03*, *04*, *05*, *06*, *08*, *09*, *12* and *16* were down-regulated in short-term- and/or long-term- alkaline treated leaves (Figure 5). These remaining *BvGATA* genes (six out of 16) were not differentially expressed in leaves under any alkaline treatments (Figure 5).

Under the salt treatment, only seven (out ò 16) *BvGATA* genes were found to be responsive in roots. Particularly, three (*BvGATA03*, *11* and *16*) and three (*BvGATA02*, *06* and *14*) genes were up- and down-regulated in salt-treated roots, respectively (Figure 5). Interestingly, *BvGATA04* was induced (~4.00-fold) and reduced (~-15.26-fold) in one-day and seven-day salt-treated root tissues (Figure 5). The expression levels of nine remaining *BvGATA* genes were not significantly changed in roots under salt treatments (Figure 5).

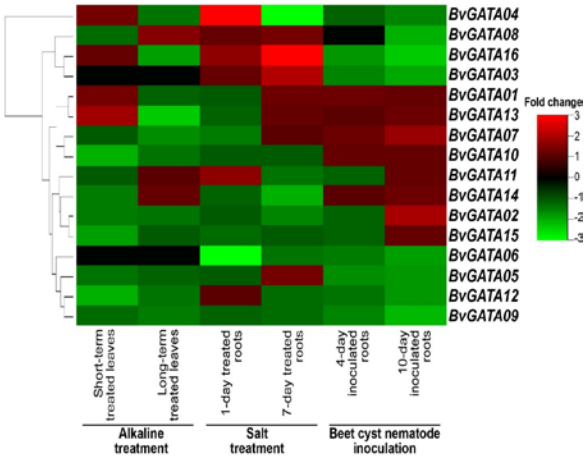

Figure 5. Expression patterns of the *BvGATA* gene family under abiotic and biotic stress

Next, our re-analysis of the microarray related to the nematode inoculation in sugar beet's roots indicated that seven (out of 16) *BvGATA* genes were down-regulated in tested tissues (Figure 5). Particularly, three genes, like *BvGATA10*, *12* and *15* were reduced in roots under four days of nematode inoculation, while three genes, including *BvGATA07*, *08* and *16* were reduced in roots under 10 days of nematode treatment (Figure 5). Interestingly, *BvGATA13* were reduced in roots under both four and 10 days of nematode inoculation (Figure 5). Our re-analysis revealed that the *BvGATA* genes were differentially expressed in major organs/tissues under treatments, which may suggest their roles during growth and development processes. In the further studies, these *BvGATA* genes would be used for functional characterization via CRISPR/Cas system to obtain the stress- resistant sugar beet lines.

4. Conclusions

In this present study, we comprehensively identify and characterize 16 BvGATA TFs in sugar beet. By using various tools, the sizes, weights, pI and AI scores of the BvGATA TFs were greatly variable, whereas all these proteins were hydrophilic. We predicted that the BvGATA TFs were localized in the cytoplasm and nucleus. Our phylogeny analysis demonstrated that the BvGATA TFs could be categorized into four distinct groups, with a majority of *BvGATA* genes contained two and three exons. We also re-analyzed the available microarray datasets and indicated that the *BvGATA* genes were significantly changed in major organs/tissues under various treatments. Taken together, our study could provide valuable information and candidate *BvGATA* genes for further functional characterization of the BvGATA TFs in sugar beet plants.

Acknowledgments

This work was funded by the fundamental research program of Hung Vuong University under the project grant No. 16/2022/HĐKH (HV16.2022).

Author Contributions

Conceptualization: P.B.C., H.D.C., T.M.L., Data collection: H.V.L, T.T.H.T., Q.T.N, T.T.N., Q.T.N.L., T.X.Q.V., T.M.L., H.G.D., Guidance of data analysis: P.B.C., D.H.C., V.H.L., T.M.L, Manuscript writing: P.B.C, D.H.C. All authors discussed the results and contributed to the final manuscript.

References

Baillo EH, Kimotho RN, Zhang Z and Xu P. 2019. Transcription Factors Associated with Abiotic and Biotic Stress Tolerance and Their Potential for Crops Improvement. *Genes*, **10(10)**: doi:10.3390/genes10100771

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S and Soboleva A. 2013. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.*, **41(D1)**: D991-D995. doi:10.1093/nar/gks1193

Behringer C and Schwechheimer C. 2015. B-GATA transcription factors–insights into their structure, regulation, and role in plant development. *Front Plant Sci.*, **6**: 90. doi: 10.3389/fpls.2015.00090

Briesemeister S, Rahnenfuhrer J and Kohlbacher O. 2010. YLoc--an interpretable web server for predicting subcellular localization. *Nucleic Acids Res.*, 38(Web Server issue): W497-502. doi:10.1093/nar/gkq477

Chen H, Shao H, Li K, Zhang D, Fan S, Li Y and Han M. 2017. Genome-wide identification, evolution, and expression analysis of GATA transcription factors in apple (*Malus x domestica* Borkh.). *Gene*, **627**: 460-472. doi:10.1016/j.gene.2017.06.049

Chen L, Zhu Y, Hu Z and Wu S. 2021. Beetroot as a functional food with huge health benefits: Antioxidant, antitumor, physical function, and chronic metabolomics activity. *Food sci nutr.*, **9(11)**: 6406-6420. doi:10.1002/fsn3.2577

Chhikara N, Kushwaha K, Sharma P, Gat Y and Panghal A. 2019. Bioactive compounds of beetroot and utilization in food processing industry: A critical review. *Food Chem.*, **272**: 192-200. doi:10.1016/j.foodchem.2018.08.022

Chu HD, Nguyen KH, Watanabe Y, Le DT, Pham TLT, Mochida K and Tran LP. 2018. Identification, structural characterization and gene expression analysis of members of the nuclear factor-Y family in chickpea (*Cicer arietinum* L.) under dehydration and abscisic acid treatments. *Int J Mol Sci.*, **19(11)**: E3290. doi:10.3390/ijms19113290

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M and Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21(18)**: 3674-3676. doi:10.1093/bioinformatics/bti610

Dohm JC, Minoche AE., Holtgräwe D, Capella-Gutiérrez S, Zakrzewski F, Tafer H, Rupp O, Sörensen TR., Stracke R, Reinhardt R, Goesmann A, Kraft T, Schulz B, Stadler PF, Schmidt T, Gabaldón T, Lehrach H, Weisshaar B and Himmelbauer H. 2014. The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature*, **505(7484)**: 546-549. doi:10.1038/nature12817

Evans E and Messerschmidt U. 2017. Review: Sugar beets as a substitute for grain for lactating dairy cattle. *J Anim Sci Biotechnol.*, **8**: 25. doi:10.1186/s40104-017-0154-8

Feng X, Yu Q, Zeng J, He X and Liu W. 2022. Genome-wide identification and characterization of GATA family genes in wheat. *BMC Plant Biol.*, **22(1)**: 372. doi:10.1186/s12870-022-03733-3

Fujita Y, Fujita M, Shinozaki K and Yamaguchi-Shinozaki K. 2011. ABA-mediated transcriptional regulation in response to osmotic stress in plants. *J Plant Res.*, **124(4)**: 509-525. doi:10.1007/s10265-011-0412-3

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD and Bairoch A. 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.*, **31(13)**: 3784-3788. doi:10.1093/nar/gkg563

Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD and Bairoch A. 2005. Protein identification and analysis tools on the ExPASy server. In: Walker JM (Eds.) The Proteomics Protocols Handbook. Springer Protocols Handbooks. Humana Press. pp. 571-607. doi: 10.1385/1-59259-890-0:571.

Ghaemi R, Pourjam E, Safaie N, Verstraeten B, Mahmoudi SB, Mehrabi R, De Meyer T and Kyndt T. 2020. Molecular insights into the compatible and incompatible interactions between sugar beet and the beet cyst nematode. *BMC Plant Biol.*, **20(1)**: 483. doi:10.1186/s12870-020-02706-8

Goodin MM. 2018. Chapter Six - Protein Localization and Interaction Studies in Plants: Toward Defining Complete Proteomes by Visualization. In Kielian M, Mettenleiter TC and Roossinck MJ (Eds.), *Adv Virus Res.* **100**, Academic Press. pp. 117-144.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N and Rokhsar DS. 2012. Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.*, **40(Database issue)**: D1178-D1186. doi:10.1093/nar/gkr944

Hall TA. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.*, **41**: 95-98.

Hu B, Jin J, Guo AY, Zhang H, Luo J and Gao G. 2015. GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics*, **31(8)**: 1296-1297. doi:10.1093/bioinformatics/btu817

Jin J, Tian F, Yang D-C, Meng Y-Q, Kong L, Luo J and Gao G. 2017. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.*, **45(D1)**: D1040–D1045. doi:10.1093/nar/gkw982

Kim M, Xi H, Park S, Yun Y and Park J. 2021. Genome-wide comparative analyses of GATA transcription factors among seven Populus genomes. *Scientific Reports*, **11(1)**: 1-15.

Kumar S, Stecher G and Tamura K. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evo.l*, **33(7)**: 1870-1874. doi:10.1093/molbev/msw054

La HV, Chu HD, Ha QT, Tran TTH, Tong VH., Tran VT, Le TNQ, Bui THT and Cao PB. 2022. SWEET Gene Family in Sugar Beet (*Beta vulgaris*): Genome-Wide Survey, Phylogeny and Expression Analysis. *Pak J biol sci.*, **25(5)**: 387-395.

Larkin MA, Blackshields G and Brown NP. 2007. Clustal W and clustal X version 2.0. *Bioinformatics*, **23**. doi:10.1093/bioinformatics/btm404

Liu L, Wang B, Liu D, Zou C, Wu P, Wang Z, Wang Y and Li C. 2020. Transcriptomic and metabolomic analyses reveal mechanisms of adaptation to salinity in which carbon and nitrogen metabolism is altered in sugar beet roots. *BMC Plant Biol.*, **20(1)**: 138. doi:10.1186/s12870-020-02349-9

Mall AK, Misra V, Santeshwari, Pathak AD and Srivastava S. 2021. Sugar Beet Cultivation in India: Prospects for Bio-Ethanol Production and Value-Added Co-Products. *Sugar tech.,* **23(6)**: 1218-1234. doi:10.1007/s12355-021-01007-0

Manzoor MA, Sabir IA, Shah IH, Wang H, Yu Z, Rasool F, Mazhar MZ, Younas S, Abdullah M and Cai Y. 2021. Comprehensive Comparative Analysis of the GATA Transcription Factors in Four Rosaceae Species and Phytohormonal Response in Chinese Pear (*Pyrus bretschneideri*) Fruit. *Int J Mol Sci.,* **22(22)**: 12492. doi:10.3390/ijms222212492

Mirmiran P, Houshialsadat Z, Gaeini Z, Bahadoran Z and Azizi F. 2020. Functional properties of beetroot (Beta vulgaris) in management of cardio-metabolic diseases. *Nutr Metab (Lond),* **17**: 3. doi:10.1186/s12986-019-0421-0

Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD and Bateman A. 2021. Pfam: The protein families database in 2021. *Nucleic Acids Res.,* **49(D1)**: D412-D419. doi:10.1093/nar/gkaa913

Monteiro F, Frese L, Castro S, Duarte MC, Paulo OS, Loureiro J and Romeiras MM. 2018. Genetic and Genomic Tools to Asssist Sugar Beet Improvement: The Value of the Crop Wild Relatives. *Front Plant Sci.,* **9**: 74. doi:10.3389/fpls.2018.00074

Niu L, Chu HD, Tran CD, Nguyen KH, Pham HX, Le DT, Li W, Wang W, Le TD and Tran L-SP. 2020. The GATA gene family in chickpea: Structure analysis and transcriptional responses to abscisic acid and dehydration treatments revealed potential genes involved in drought adaptation. *J Plant Growth Regul.,* **39(4)**: 1647-1660. doi:10.1007/s00344-020-10201-5

Pavlečić M, Rezić T, Šantek MI, Horvat P and Šantek B. 2017. Bioethanol production from raw sugar beet cossettes in horizontal rotating tubular bioreactor. *Bioprocess Biosyst Eng.,* **40(11)**: 1679-1688. doi:10.1007/s00449-017-1823-x

Peng W, Li W, Song N, Tang Z, Liu J, Wang Y, Pan S, Dai L and Wang B. 2021. Genome-Wide Characterization, Evolution, and Expression Profile Analysis of GATA Transcription Factors in *Brachypodium distachyon*. *Int J Mol Sci.,* **22(4)**: 2026. doi:10.3390/ijms22042026

Reyes JC, Muro-Pastor MI and Florencio FJ. 2004. The GATA family of transcription factors in *Arabidopsis* and rice. *Plant Physiol.,* **134(4)**: 1718-1732. doi:10.1104/pp.103.037788

Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE and Sanchez-Gracia A. 2017. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol Biol Evol.,* **34(12)**: 3299-3302. doi:10.1093/molbev/msx248

Schwechheimer C, Schroder PM and Blaby-Haas CE. 2022. Plant GATA Factors: Their biology, phylogeny, and phylogenomics. *Annu Rev Plant Biol.,* **73**: 123-148. doi:10.1146/annurev-arplant-072221-092913

Teakle GR., Manfield IW, Graham JF and Gilmartin PM. 2002. *Arabidopsis thaliana* GATA factors: organisation, expression and DNA-binding characteristics. *Plant Mol Biol.,* **50(1)**: 43-57. doi:10.1023/a:1016062325584

Yu C, Li N, Yin Y, Wang F, Gao S, Jiao C and Yao M. 2021. Genome-wide identification and function characterization of GATA transcription factors during development and in response to abiotic stresses and hormone treatments in pepper. *J Appl Genet.,* **62(2)**: 265-280. doi:10.1007/s13353-021-00618-3

Yu R, Chang Y, Chen H, Feng J, Wang H, Tian T, Song Y and Gao G. 2021. Genome-wide identification of the GATA gene family in potato (*Solanum tuberosum* L.) and expression analysis. *J Plant Biotech Biochem.,* **31(1)**: 37-48. doi:10.1007/s13562-021-00652-6

Zhang C, Hou Y, Hao Q, Chen H, Chen L, Yuan S, Shan Z, Zhang X, Yang Z, Qiu D, Zhou X and Huang W. 2015. Genome-wide survey of the soybean GATA transcription factor gene family and expression analysis under low nitrogen stress. *PLoS One,* **10(4)**: e0125174. doi:10.1371/journal.pone.0125174

Zhang C, Huang Y, Xiao Z, Yang H, Hao Q, Yuan S, Chen H, Chen L, Chen S, Zhou X and Zhou X. 2020. A GATA transcription factor from soybean (*Glycine max*) regulates chlorophyll biosynthesis and suppresses growth in the transgenic *Arabidopsis thaliana*. *Plants (Basel),* **9(8)**: 1036. doi:10.3390/plants9081036

Zhang K, Jia L, Yang D, Hu Y, Njogu MK and Wang P. 2021. Genome-Wide Identification, Phylogenetic and Expression Pattern Analysis of GATA Family Genes in Cucumber (*Cucumis sativus* L.). *Plants (Basel, Switzerland).* **10(8)**: 1626. doi:10.3390/plants10081626

Zhang Y, Nan J and Yu B. 2016. OMICS Technologies and Applications in Sugar Beet. *Front Plant Sci.,* **7**: 900-900. doi:10.3389/fpls.2016.00900

Zhang Z, Ren C, Zou L, Wang Y, Li S and Liang Z. 2018. Characterization of the GATA gene family in *Vitis vinifera*: genome-wide analysis, expression profiles, and involvement in light and phytohormone response. *Genome,* **61(10)**: 713-723. doi:10.1139/gen-2018-0042

Zhang Z, Zou X, Huang Z, Fan S, Qun G, Liu A, Gong J, Li J, Gong W, Shi Y, Fan L, Zhang Z, Liu R, Jiang X, Lei K, Shang H, Xu A and Yuan Y. 2019. Genome-wide identification and analysis of the evolution and expression patterns of the GATA transcription factors in three species of *Gossypium* genus. *Gene,* **680**: 72-83. doi:10.1016/j.gene.2018.09.039

Zou C, Wang Y, Wang B, Liu D, Liu L, Gai Z and Li C. 2020. Long non-coding RNAs in the alkaline stress response in sugar beet (*Beta vulgaris* L.). *BMC Plant Biol.,* **20(1)**: 227. doi:10.1186/s12870-020-02437-w**.**